

The Effect of Choice of Prompts on Syntactic Complexity, Grammatical Accuracy, and Lexical Diversity in L2 Argumentative Writing Essays

Rajab Esfandiari¹  Nilofar Sadeghian²

¹ *Corresponding author*, Associate Professor, Department of English Language, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran, Email: esfandiari@hum.ikiu.ac.ir

² M.A., Department of English Language, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran, Email: niloo_s70@yahoo.com

Abstract

Previous research has shown the differential effects of task-related prompts on syntactic complexity, grammatical accuracy, and lexical diversity when L2 learners use writing prompts to produce a piece of writing. However, the extent to which the freedom in the selection of prompts affects these linguistic facets in L2 argumentative essays is still unknown. The present study, therefore, was designed to investigate differences between syntactic complexity, grammatical accuracy, and lexical diversity in argumentative essays. Seventy-one upper-intermediate male and female Iranian English-as-a-foreign language (EFL) learners from Imam Khomeini International University in Qazvin and Safir Plus institute in Tehran participated in this study. Five-paragraph essays were used to collect data, and SPSS (version 25) was used to analyze the data. Results of multivariate analysis of variance (MANOVA) showed statistically significant differences between syntactic complexity, grammatical accuracy, and lexical diversity. Results from post-hoc analyses revealed statistically significant differences between lexical diversity and grammatical accuracy as well as lexical diversity and syntactic complexity, but no statistically significant differences were found between grammatical accuracy and syntactic complexity in argumentative essays.

Keywords: argumentative writing, grammatical accuracy, lexical diversity, syntactic complexity

1. Introduction

Language learners advance at different paces, as the tasks they encounter increase in complexity, accuracy, and diversity. This variation is often attributed to language learners' limited cognitive capacities, although their effective performance in tasks largely relies on their progress in language learning (Robinson, 2015). Among four language skills, writing seems to be the most complicated one, as it draws on the cognitive domain, including learning and comprehension of recent information (Defazio et al., 2010). Studies focusing on L2 writing and the effect of task design features such as complexity, accuracy, and lexical diversity have yielded mixed results. For example, personal information exchange tasks, familiar information, interactive tasks, and post-task conditions increase accuracy; however, pre-task planning and narrative tasks produce greater syntactic complexity, but lower grammatical accuracy (e.g., Frear & Bitchener, 2015; Allaw, & McDonough, 2019; Kessler et al., 2021; Martin-Beltrán et al., 2020).

The effect of tasks on L2 writing has been studied from two psycholinguistic perspectives: Cognition hypothesis (Robinson, 2011) and trade-off hypothesis (Skehan & Foster, 2001). Cognition hypothesis claims that complexity, accuracy, and fluency have an impact on learning, as the act of learning utilizes various levels of cognitive demand on L2 learners' cognitive sources (Skehan, 2009). However, Skehan and Foster's (2001) trade-off hypothesis posits that when the task gets more complex, the learner's focus on form decreases.

Previous studies on syntactic complexity, grammatical accuracy, and lexical diversity have yielded some interesting, though inconsistent, results. Some studies have investigated the effect of syntactic complexity on the quality of argumentative essay (e.g., Rahimi, 2018), while other studies have analyzed syntactic complexity in different genres of writing such as descriptive essays (Beers & Nagy, 2011). Other studies have examined syntactic complexity in expository writing among students with and without language learning disabilities (Koutsoftas & Gray, 2012). Other researchers have analyzed grammatical accuracy in argumentative essays (Tavakoli & Rezazadeh, 2014). However, other researchers have analyzed the effect of tasks and prompts on grammatical accuracy across expository and descriptive essays (Way et al., 2000).

Other researchers have investigated the relationship of lexical diversity with other aspects of language. For example, Dilmaghani and Sadeghi (2013) investigated the relationship between lexical diversity and two genres of writing—argumentative and descriptive—in Iranian EFL learners' writing and found a positive correlation between text types and lexical diversity.

Although previous research on syntactic complexity, grammatical accuracy, and lexical diversity has helped researchers better understand the dynamics and processes of these three constructs across different text types, modes of language, settings, and learner groups and has shed light on the theoretical underpinnings of these concepts, one major shortcoming of previous studies relates to the assertion that researchers have used a combination of variables to analyze complexity, accuracy, and diversity, resulting in confounding findings (see Ahmadian & Long, 2022 for the variety of variables examined so far). In the present study, we, however, examined syntactic complexity, grammatical accuracy, and lexical diversity on

upper-intermediate language learners in argumentative essays to control two major variables on our findings: language proficiency and text types. We also asked language learners to feel free to choose from a list of prompts to write their argumentative essays, a point not addressed in previous studies, as explained below.

Prior research has shown that several factors tend to affect syntactic complexity, grammatical accuracy, lexical diversity across different text types, including discourse mode (Liao, 2020), purpose of writing (Rahimi & Zhang, 2021), topic structure and wording (Ahmadian & Mansouri, 2020), rhetorical specification (Allaw, & McDonough, 2021), text length (Nasseri & Thompson, 2021), different genre types (Zhang & Lu, 2021), and scoring procedures (Weigle, 2018). Previous studies focusing on writing prompts across text types concerning complexity, accuracy, and diversity have also reported on the following results.

Prompts as themes tend to exert no effect on the scores and quality of argumentation in graduate EFL argumentative texts (Jalilifar et al., 2017). Self-selected versus teacher selected prompts yield in more complex and accurate undergraduate writing across three proficiency levels (beginning, intermediate, and advanced) (Meihami & Saadat, 2019). More implicit prompts result in less accurate production, higher syntactic complexity, and more varied modes of argumentation (Shi et al., 2020). Last but not least, recalling prompts vs imagining prompts leads to more complex syntax, and more lexical diversity, but less accurate sentences in argumentative texts (Cho, 2019). However, as Hamp-Lyons (2011) noted, questions such as “should writers be offered a choice of tasks on a writing test? ... have not received definitive answers” (p. 5). It was to this basic question we turned our attention and set out to investigate the choice of prompts and their effect on syntactic complexity, grammatical accuracy, and lexical diversity in Iranian EFL learners’ writing because of the following two major motives. Complexification, diversity, and accuracy, as we explain below, appear to be indicators of L2 writing proficiency and can offer insights into practical methods about how to develop writing proficiency in EFL settings.

The first reason is related to the significance of writing prompts, because they help students to have the best possible option to demonstrate accurately their true level of writing and provide “the stimulus for the students to respond to (Kroll & Reid, 1994, p. 231). Prompts are also critical task related features which affect students’ writing performance, resulting in complex, accurate, and diverse language (Shi et al., 2020).

Writing prompts are contextual factors which provide direction for writing, frame writing, function as a base for assessing students’ written products, and “provide space for meaning-making in L2 writing as they elicit diverse learner responses, yet in a relatively framed way” (Cho, 2019, p. 578). The second main motivating factor for undertaking this study was because syntactic complexity, grammatical accuracy, and lexical diversity have received considerable attention from researchers because they have proved useful and reliable measures of second language performance, development, and proficiency. For example, the proficiency level of learners is measured through the ability to produce lexically diverse, syntactically complex, and grammatically accurate sentences in writing (Housen et al., 2012). Therefore, the present researchers formulated the following research question to focus their study:

What is the effect of choice of prompts by language learners on syntactic complexity, grammatical accuracy, and lexical diversity in argumentative essays?

2. Literature Review

2.1. Syntactic Complexity, Grammatical Accuracy, and Lexical Diversity in L2 Development

Complexity is one of the most challenging constructs, as it is multi-faceted. Ellis and Barkhuizen (2005) proposed a list of aspects of complexity, including lexical, interactional, propositional, and grammatical or syntactic complexity. Syntactic complexity is defined as the ability to use more complex sentence structures (Housen & Kuiken, 2009). Robinson (2011) claims that increasing task complexity will increase syntactic complexity. Similarly, Ruiz-Funes (2015) argues that complex task performance results in higher syntactic complexity, irrespective of learners' proficiency.

There are at least two contemporary branches of SLA research which capitalize on the concept of syntactic complexity. The first strand views syntactic complexity as an independent variable, in which its influence on some aspects of L2 performance, or L2 proficiency, is investigated (Housen et al., 2012). Several studies have looked at the impact of complexity on its teachability, or on the effectiveness, of different kinds of instruction (e.g., Spada & Tomita, 2010). The second strand views syntactic complexity as a dependent variable and as a basic indicator of L2 performance and L2 proficiency (Bulté & Housen, 2012).

Most studies which have used syntactic complexity either as an independent or a dependent variable have yielded mixed or contradictory results (e.g., Robinson, 2009; Skehan, 2009). These inconsistent results can be accounted for by the way in which complexity is defined and measured (Housen et al., 2012). Researchers have offered various measures of syntactic complexity, which include length, subordination, and coordination, each with its own proponents and opponents (Bayazidi et al., 2019), although recently NLP has provided better orientation towards syntactic complexity

Accuracy is probably the oldest and the most consistent construct, although accuracy per se is not an indicator of interlanguage development (Pallotti, 2009). Foster and Skehan (1999) defined grammatical accuracy as the ability to avoid errors in performance. However, Pallotti (2009) refers to grammatical accuracy as "the degree of conformity to certain norms" (p. 592). There is a debate as to how accuracy should be defined (Towell, 2012). For example, much earlier, Selinker (1972) pointed out that the results could be juxtaposed with the collective performance of native speakers, although not all researchers agree with her.

Several researchers have attempted to investigate the effect of some factors of writing on grammatical accuracy. Kuiken and Vedder (2012) proposed that as students were asked to perform more complex tasks, their grammatical accuracy would drop. Also, Rahmihi's (2018) work on the effects of increasing the degree of reasoning and the number of elements on L2 argumentative writing confirmed Kuiken and Vedder's findings. However, these findings contradict the results of other studies as other researchers concluded that when the task complexity increased, the grammatical accuracy would increase as well (Ishikawa, 2007;

Kuiken & Vedder, 2011). Wigglesworth and Storch (2009) are other researchers who compared pair and individual writing and their effect on grammatical accuracy in their study. The results showed that writing in pairs positively affected grammatical accuracy. However, individual writing affected grammatical accuracy negatively.

Researchers have suggested various measures of grammatical accuracy. These measures are based on the number of error-free T-units (Polio, 2001), error-free T-units per T-unit (Scott & Tucker, 1974 as cited in Jiang, 2013), and errors per T-unit (Polio, 2001). T-unit was defined by Hunt (1965 as cited in Foster et al., 2000) as “one main clause plus whatever subordinate clauses happen to be attached to or embedded within it” (p. 735). Several researchers have used the number of error-free T-units (EFT/T) as a measure of grammatical accuracy (e.g., Rutherford, 2001). Wolfe-Quintero et al. (1998) claimed, in their meta-analysis of studies of L2 writing, that the EFT/T measure is the most effective measure of written accuracy. Other researchers have used the number of error-free T-units per T-unit to measure accuracy. For example, Elder and Iwashita (2005) used this measure as one of the measures of accuracy. Finally, some researchers measured grammatical accuracy by measuring the number of errors per T-unit (e.g., Guara-Tavares, 2008).

Lexical diversity (LD) has been considered a predictor of learners' language proficiency. MacCarthy and Jarvis (2010) defined lexical diversity as “the range of different words used in a text, with a greater range indicating a higher diversity (p. 381). LD provides a measure of the proportion of lexical items (i.e., nouns, verbs, adjectives, and some adverbs) in the text (Johansson, 2009). Research has shown a strong relationship between high-quality academic writing, lexical diversity, and the use of academic and lower frequency words in samples produced by NSs (native speakers) of English and ESL (English as a second language) students at various educational levels (McNamara et al., 2013).

Measuring lexical diversity is as complicated as the concept itself is. Type-token ratio (TTR) is the best known measure of LD, and it shows the ratio of different words (types) to total words (tokens) in a given language sample. TTR is usually credited with Templin (1957 as cited in Treffers-Daller et al., 2016). The key problem with TTR is that it is sensitive to sample size, which means TTR varies inversely depending on sample size; in other words, the shorter the sample, the greater the value of TTR.

Other measures of LD, therefore, have been developed recently. These include Malvern et al.'s (2004) D value, McCarthy and Jarvis's (2007) HD-D index, and McCarthy's (2005) MTLTD. A key problem for D, HD-D, and MTLTD is text length. While MTLTD decreases with text length, HD-D and D increase with it. Further studies are still needed for developing LD measures, which are not dependent on text length. All things considered, it is sometimes more successful to measure LD by TTR rather than using a complex formula such as D.

2.2. Text Types in Academic Writing

Academic writing differs from other types of writing. Many English speakers believe that the language of academic writing differs considerably not only from everyday speech but also from most other registers

of English (Biber & Gray, 2016). Therefore, the common stereotype is that academic writing is rather complex (Ädel & Erman, 2012). The first reason of these attitudes is because of using rare and obscure words and phrases (Biber, 2006). The second reason is that grammatical features will lead to the general perception that academic writing is more complex and explicit in meaning than speech production and even other registers in writing (Biber & Gray, 2016) and the last reason is that in contrast to spoken grammar, written grammar employs longer and more complex clauses (Foster et. al., 2000).

In academic writing, three main genres are commonly used to help writers to communicate ideas to readers for various purposes. These three important genres are argumentative, narrative/descriptive, and expository. The purpose of argumentative writing is to present a position to convince or persuade a given group of audience to understand the other side of the argument or to support a new belief or idea (Newell et al., 2011), while the goal of descriptive writing is to describe a thing, a place, or a person in a way that a picture is formed in the reader's mind (McCarthy, 1998). In contrast to argumentative and descriptive writing, the aim of expository writing is to explain and analyze information by presenting an idea (Hall et al., 2005).

Writing in different styles can be a linguistically demanding task. Genre researchers have shown that writing in different genres requires different linguistic demands (Nippold, 2004) and different cognitive task loads (Kamberelis & Bovino, 1999). Writing argumentative and expository texts is cognitively more demanding than writing a narrative or descriptive text (Weigle, 2002). Brand-Gruwel et al. (2005) believe that writing an argumentative essay is a complex cognitive process because the reader's expectations, the writer's goals, the rhetorical patterns, and the contextual situations are engaged.

The argumentative genre has been one of the most favored genres in writing assessment. However, argumentative writing is a complex activity because the writer will be engaged with a controversial issue, giving reasons and supporting evidence to convince the reader to accept his or her position (Bean, 2011). In this respect, Abdollahzadeh and Amini Farsani (2017) examined the argumentative behavior of Iranian graduate EFL learners who wrote 150 argumentative essays. The results of their study showed that the learners tended to organize their essays deductively than inductively. "Data" and "claim" were the most frequently used argument elements; by contrast, secondary elements such as "counterargument claim", "counterargument data", "rebuttal claim", and "rebuttal data" were the least produced argument elements. The findings further indicated that overall writing quality correlated positively with the uses of, i.e., claim, data, and rebuttal claims. Abdollahzadeh and Amini Farsani concluded that "even for advanced language learners good surface structure cannot necessarily guarantee well thought-out logical structure" (p. 641).

Narrative, or descriptive genre, is another text type which is considered by teachers and learners. Narrative, or descriptive writing, includes biographies, tales, fiction, and historical events (Grabe, 2002). Findings have shown that the vocabulary used in narrative writing contains more personal pronouns and words with more sensory images (Berman & Nir-Sagiv, 2007). In comparison to argumentative writing and expository writing, narrative genres have shorter clauses (Malvern et al., 2004), fewer complex noun phrases (Ravid & Berman, 2010), fewer relative and adverbial clauses (Scott & Windsor, 2000), and less passive

voice (Reilly et al., 2005).

Planning and writing narrative essays do not need a high cognitive effort. Beauvais et al. (2011) found that students spent less time planning to write a narrative text because it required a less complex and sophisticated knowledge-transforming strategy. Thus, the prewriting stage for narrative essays required less cognitive effort (Jeong, 2017). The expository writing covers a wide range of non-narrative texts such as persuasive, compare and contrast, and procedural texts (Beers & Naggy, 2011). Researchers have found that the expository texts display a more advanced lexis (Berman & Nir-Sagiv, 2004) and expository genres have longer clauses (Malvern et al., 2004), more complex noun phrases (Ravid & Berman, 2010), more relative and adverbial clauses (Scott & Windsor, 2000), and more passive voice (Reilly et al., 2005). Writing an expository essay is a more demanding task than writing a narrative essay. In comparison to narrative genres, Beauvais et al. (2011) believe that learners spend more time on planning an expository text because it needs a more complex strategy. Unfamiliar elements, the presence of abstract, complex relations, high informational density, processing of more detailed information, and considerable variations in expository texts such as different text structures are the additional factors which, as Jeong (2017) noted, render expository texts to be cognitively more demanding.

3. Methodology

3.1. Participants

Both male and female upper-intermediate Iranian language learners attending language classes at Safir Plus and Imam Khomeini International University participated in the present study. A proficiency test was used to select only upper-intermediate language learners, because previous research has shown argumentative essays are difficult for low-proficiency students to produce (e.g., Allaw & McDonough, 2019). The participants included our analysis were 60 upper-intermediate language learners who were conveniently selected from 200 learners from the above language institute and university.

3.2. Instruments

Oxford Quick Placement Test (OQPT) (Version 2) (UCLES, 2001) was used to measure the participants' proficiency level. The placement test is divided into two parts: Part 1 (Questions 1-40) and part 2 (Question 41-60). Part 1 was given to all the students, and part two started when language learners had answered questions in part one without any problems. The participants had 30 minutes to answer all the questions. Geranpayeh (2006) argued that OQPT, which is a standardized English proficiency test, had been pretested and validated by about 6,000 students in about 60 countries. According to Allan (2004), the developer of the test, OQPT has been calibrated against the proficiency levels based on the Common European Framework of Reference for Languages (CEF), the Cambridge ESOL Examinations, and other major international examinations such as TOEFL. The cut-off points for proficiency levels set by Allan

(2004) was considered by several researchers (e.g., Jabbari, 2014; Rebarber et al., 2007) as reliable indicators that would signal language proficiency levels. Based on the level that the scores of the Oxford Quick Placement test determines, students whose placement test scores were less than 40 and more than 47 were excluded as they are not in upper-intermediate level to be included in this study.

Five-paragraph argumentative essays were also used to collect writing data. Language learners were presented with seven prompts to develop their essays. They felt free to choose from this list topics because previous research has shown when language learners are forced to choose a topic, it affects the writing they produce. To reduce the effect of external topic pressure, we collected 20 prompts from TOEFL and IELTS practice tests commercially available on the market, selected seven of them, and asked language learners to choose the one he or she felt comfortable with to write about.

Frequency and familiarity were two criteria we used to select seven out of 20 prompts for our study. The 20 prompts were piloted on language learners in another writing class prior to the start of the actual study and language learners were asked to select the most interesting prompts they would like to write about. We tallied the number of times each prompt was selected by these students, and chose the most frequently endorsed prompts by these students, the result of which includes seven prompts in Appendix. To ensure all the students were familiar with these seven prompts, we invited them to choose one of those which they had enough knowledge to write.

3.3. Procedure

The following steps were used in this study. At first, Oxford Quick Placement Test (OQPT) (Version 2) was given to the participants to determine their proficiency level. Based on the criteria of Oxford Quick Placement Test, to determine the learners' level of proficiency, students whose scores were lower than 40 and higher than 47 were excluded. Then, the researchers gave them each one of those seven prompts. The researchers, then, asked them to produce a timed argumentative essay between 250 to 300 words in 40 minutes. Finally, the essays were collected and elements such as syntactic complexity, accuracy, and lexical diversity were measured with the following steps.

Three measures of syntactic complexity were employed in this study: (a) T-units, defined as "one main clause plus whatever subordinate clauses happens to be attached to or embedded within it" (Hunt's, 1966, p.735). The *t*-unit was adopted due to the monologic nature of writing as a form of language production; (b) The mean length of clauses, which was proposed by Scott (1988) as a "measure of sub-clausal complexity which is typically achieved via grammatical metaphor such as phrasal elaboration and nominalization; and, (c) subordination (the ratio of subordinated clauses to the total number of clauses) as a measure of clausal complexity, along with phrasal coordination" (Rahimi, 2018). The grammatical accuracy of the learners' writing was measured by computing the ratio of error-free *t*-units to the total number of *t*-units. An error-free *t*-unit is defined as a *t*-unit which is free of any lexical, grammatical, and morphological errors. All errors in punctuation and spelling were ignored.

To measure LD, the value of D (Malvern & Richard, 2002) was computed by using Text Inspector (Bax, 2018; Williams, 2018). Text Inspector is the professional web tool for analyzing texts which was developed by Bax (2018), and it allows users to analyze writing by providing data on many aspects of a text's lexis. All users need to do is input their text into a box on the webpage and a comprehensive analysis is produced. This study used values of D as it can be calculated with high reliability even for short texts (Chipere et al., 2004).

We have focused on lexical diversity, grammatical accuracy, and syntactic complexity in argumentative essays because of the following reasons. Previous research (e.g., Kessler et al., 2021) has shown that argumentative essay writers usually employ a wide range of lexical items, complex syntactic structures, and well-formed grammatical constructions to convince readers of the point(s) they make. However, lexical diversity, grammatical accuracy, and syntactic complexity may not be given an equal weight in argumentative texts, and differences exist between these constructs, when other variables such as topic familiarity are at play. One such variable, which we examined in this study, was the effect of prompts the differences between lexical diversity, grammatical accuracy, and syntactic complexity.

3.4. Data Analysis

The data were analyzed by Statistical Package for Social Sciences (SPSS) in the following steps. Before analyzing the data, the assumptions of the statistical test we used in this study were checked for assumptions. Multivariate analysis of variance (MANOVA) was used to estimate the effect of the choice of prompts, or tasks, on each element under study (i.e., syntactic complexity, accuracy, and lexical diversity). An alpha level of $p < .05$ was regarded acceptable. To measure effect size, partial eta squared (Gray & Kinnear, 2012) was computed for MANOVA; according to the guidelines (Gray & Kinnear, 2012), f values of [0.10, 0.25), [0.25, 0.40), and ≥ 0.40 are considered small, medium, and large, respectively.

4. Results

4.1. Testing Assumptions

To study how choice of prompts of three different types of essays affects grammatical accuracy, lexical diversity, and syntactic complexity, multivariate analysis of variance (MANOVA) test was used. Before discussing the results, it should be noted that the assumptions of lack of univariate and multivariate outliers, univariate and multivariate normality, linearity, lack of singularity and multicollinearity, and homogeneity of variances-covariances were checked, as reported below.

4.1.1. Lack of Univariate and Multivariate Outliers

The present data were scrutinized for any univariate and multivariate outliers. The standardized scores (Z-scores) were computed for syntactic complexity, grammatical accuracy, and lexical diversity in

order to check lack of univariate outliers. Table 1 displays the descriptive statistics for the Z-scores for these three variables. Any Z-score higher than ± 3 was considered to be an outlier at .05 level. The results indicated that the Z-score on accuracy was higher than -3. An inspection of the data revealed that ID number 55 score on the accuracy test was .10. The present researchers reran the data with and without this outlier, and because the results were similar, they decided not to remove it, as suggested in Tabachnick and Fidell (2014). The other two variables did not show any significant univariate outliers with Z-scores higher than ± 3 .

Table 1*Descriptive Statistics; Standardized Scores (Univariate Outliers)*

	N	Minimum	Maximum	Mean	Std. Deviation
Zscore (Accuracy)	60	-3.084	1.826	.000	1.00
Zscore (Complexity)	60	-1.821	2.420	.000	1.00
Zscore (Diversity)	60	-1.803	2.323	.000	1.00

Lack of multivariate outliers was probed through Mahalanobis Distances (MD) which were compared against the critical value of chi-square at .001 level for three variables, i.e., 16.26 (Tabachnick & Fidell, 2014). Table 2 displays the descriptive statistics for the MD. Since maximum MD of 11.47 was lower than 16.26, it was concluded that there were not any multivariate outliers in the data set.

Table 2*Descriptive Statistics of Mahalanobis Distances; Testing Multivariate Outliers*

	N	Minimum	Maximum	Mean	Std. Deviation
Mahalanobis Distance	60	.141	11.47	2.949	2.229

4.1.2. Lack of Univariate and Multivariate Normality

The data were then examined for univariate and multivariate normality. The Kolmogorov-Smirnov tests were computed to check univariate normality of the present data. Since all probabilities associated with the Kolmogorov-Smirnov test were higher than .05, it was concluded that the assumption of univariate normality was retained, as shown in Table 3.

Table 3*Tests of Univariate Normality*

		Kolmogorov-Smirnov		
	Group	Statistic	df	Sig.
Grammatical Accuracy	Argumentative	.098	59	.200
Syntactic Complexity	Argumentative	.121	59	.200
Lexical Diversity	Argumentative	.147	59	.200

The assumption of multivariate normality was checked through Mardia's index using Stata/IC computer programme (version. 14.2) (StataCorp, 2016) software. Table 4 displays the results of the joint skewness and kurtosis tests. The non-significant results of the test supported the multivariate normality of the data.

Table 4*Joint Skewness-Kurtosis Test of Normality*

Dependent Variables	Joint Skewness-Kurtosis	
	Chi-square	P value
Accuracy	5.97	.053
Complexity	1.63	.442
Diversity	2.37	.306

The second part of the Stat's output (Table 5) examines bivariate normality of the variables. The results indicated that the two-by-two combinations of the three tests met the normality assumption ($p > .05$).

Table 5*Bivariate Test of Normality*

Dependent Variables	Joint Skewness-Kurtosis		
	Chi-Square	df	P value
Accuracy with Complexity	8.88	4	.064
Accuracy with Diversity	8.74	4	.068
Complexity with Diversity	3	4	.558

The final output (Table 6) displayed the Mardia's multivariate skewness and kurtosis indices. Since the probabilities associated with the multivariate skewness ($p = .473$) and multivariate kurtosis ($p = .480$) were higher than .05, it was concluded that the assumption of multivariate normality was retained.

Table 6*Mardia's Test of Multivariate Normality*

Mardi's Test	Joint Skewness-Kurtosis		
	Chi-Square	Df	p-value
Multivariate Skewness	9.63	10	.473
Multivariate Kurtosis	.498	1	.480

4.1.3. Testing Lack of Singularity and Multicollinearity

The assumption of lack of singularity, i.e., perfect correlation among all variables, was examined by computing the determinant index (Table 7). The SPSS package produces determinant index as a part of its output for factor analysis. As displayed below, the determinant value of .899 was higher than .00001 (Field, 2018). Thus, it can be concluded that the present data did not suffer from singularity.

Table 7*KMO and Bartlett's Test of Sphericity*

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.493
Bartlett's Test of Sphericity	Approx. Chi-Square	5.988
	Df	3
	Sig.	.112
Determinant		.899

The assumption of multicollinearity can be checked using the SPSS software's Linear Regression command. As displayed in Table 8, two sets of statistics were produced to test the assumption of

multicollinearity. The tolerance indices were all higher than .10. Thus, it can be concluded that the assumption of multicollinearity was retained. The variance inflation factor (VIF) indices were also all lower than 10. These results also support the multicollinearity of the present data.

Table 8*Testing Assumption of Multicollinearity*

Model	Unstandardised Coefficients		Standardised Coefficients	<i>t</i>	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	4.485	3.957		1.134	.262		
Grammatical Accuracy	1.769	2.877	.087	.615	.541	.899	1.112
Syntactic Complexity	.701	5.257	.019	.133	.894	.917	1.090
Lexical Diversity	.614	2.447	.034	.251	.803	.979	1.021

4.1.4. Homogeneity of Variance-Covariance

The assumption of homogeneity of variances was checked using the Levene's test. The non-significant results of Levene's tests indicated that the assumption of homogeneity of variances was retained on grammatical accuracy ($F_{(2, 56)} = .396, p = .675$), syntactic complexity ($F_{(2, 56)} = .409, p = .666$) and lexical diversity ($F_{(2, 56)} = 1.44, p = .244$) (Table 9).

Table 9*Levene's Test of Equality of Error Variances*

	Levene Statistic	df 1	df 2	Sig.
Accuracy	Based on Mean	.332	2	.719
	Based on Median	.396	2	.675
	Based on Median and with adjusted df	.396	2	55.797
	Based on trimmed mean	.411	2	.665
Complexity	Based on Mean	.487	2	.617
	Based on Median	.409	2	.666
	Based on Median and with adjusted df	.409	2	55.930
	Based on trimmed mean	.489	2	.616
Diversity	Based on Mean	1.446	2	.244
	Based on Median	1.379	2	.260
	Based on Median and with adjusted df	1.379	2	55.476
	Based on trimmed mean	1.482	2	.236

Table 10 shows the results of the Box's test. The non-significant results of the Box's test ($M = 16.90, p = .211$) indicated that the assumption of homogeneity of covariance matrices was met. That is to say, the correlations between any two dependent variables were roughly equal across the groups.

Table 10*Box's Test of Equality of Covariance Matrices^a*

Box's M	16.906
F	1.299
df1	12
df2	15110.285
Sig.	.211

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Group

Within Subjects Design: writing

4.1.5. Assumption of Sphericity

The assumption of sphericity was also met. MANOVA requires that the difference between any two dependent variables have roughly equal variances. The non-significant results of Mauchly's sphericity test ($W=.790$, $p=.059$) (Table 11) indicated that the assumption of sphericity was retained.

Table 11*Mauchly's Test of Sphericity for Accuracy, Complexity and Vocabulary Density (Argumentative Essays)*

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Argumentative	.730	5.658	2	.059	.788	.847	.500

4.2. Investigation of the Research Question

To answer the research question (What is the effect of choice of prompts by language learners on syntactic complexity, grammatical accuracy, and lexical diversity in argumentative essays?), multivariate analysis of variance (MANOVA) was used since the normality and other assumptions were retained for the argumentative essays.

Table 12 shows the descriptive statistics for syntactic complexity, grammatical accuracy, and lexical diversity for the argumentative genre. The results indicated that argumentative essays included the highest mean on lexical diversity ($M=.834$), which was followed by grammatical accuracy ($M=.599$), and finally syntactic complexity ($M=.487$).

Table 12*Descriptive Statistics for Grammatical Accuracy, Syntactic Complexity and Lexical Diversity of Argumentative Essays*

Descriptive	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Grammatical Accuracy	.599	.048	.498	.699
Syntactic Complexity	.487	.020	.446	.529
Lexical Diversity	.834	.035	.760	.908

Table 13 shows the results of MANOVA. The results ($F_{(2, 18)}=41.92$, $p= .000$, $\eta^2= .823$ representing a large effect size) indicated that there were significant differences between the means on lexical diversity, grammatical accuracy, and syntactic complexity in argumentative essays.

Table 13*Multivariate Tests for Grammatical Accuracy, Syntactic Complexity, and Lexical Diversity of Argumentative Essays*

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's Trace	.823	41.921	2	18	.000	.823
Wilks' Lambda	.177	41.921	2	18	.000	.823
Hotelling's Trace	4.658	41.921	2	18	.000	.823
Roy's Largest Root	4.658	41.921	2	18	.000	.823

The significant results of MANOVA were followed by post-hoc comparison tests in order to compare the variables two by two (Table 14). Based on the results displayed in Table 12 and Table 13, it can be concluded that (1) a significantly higher mean is seen on lexical diversity ($M=.834$) than grammatical accuracy ($M=.599$) ($MD=.235$, $p=.001$), (2) significantly higher mean goes to lexical diversity ($M=.834$) than syntactic complexity ($M=.487$) ($MD=.346$, $p=.000$), and (3) there were not any significant differences between the means of syntactic complexity ($M=.487$) and grammatical accuracy ($M=.599$) ($MD=.111$, $p=.095$).

Table 14*Multiple Comparison Tests for Grammatical Accuracy, Syntactic Complexity, and Lexical Diversity of Argumentative Essays*

(I) Descriptive	(J) Descriptive	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Diversity	Accuracy	.235*	.058	.001	.115	.356
	Complexity	.346*	.038	.000	.266	.427
Accuracy	Complexity	.111	.063	.095	-.021	.244

*. The mean difference is significant at the .05 level.

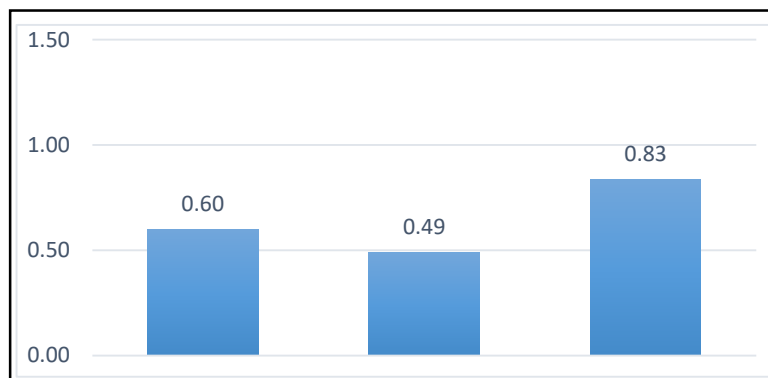
The result of between-subjects effects is demonstrated in Table 15. The average means of these three measures is statistically significant ($F_{(1, 19)}=1197.88$, $p=.000$). Figure 1 graphically shows the rounded mean values for grammatical accuracy, syntactic complexity, and lexical diversity.

Table 15*Tests of Between-Subjects Effects of Argumentative Essays*

Transformed Variable: Average							
Group	Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
Argumentative	Intercept	24.579	1	24.579	1197.880	.000	
	Error	.390	19	.021			

Figure 1

Mean values on Lexical Diversity, Syntactic Complexity, and Grammatical Accuracy in Argumentative Essays



5. Discussion

The present study was designed to explore if there were any significant differences between syntactic complexity, grammatical accuracy, and lexical diversity in argumentative essays when freedom of choice in prompts was given to the learners. Based on the results of statistical analysis, statistically significant differences were found between lexical diversity, grammatical accuracy, and syntactic complexity in argumentative essays ($p\eta^2 = .823$), considered large judging by the standards given in Cohen (1988). Results from post-hoc analyses showed statistically significant differences between lexical diversity and grammatical accuracy as well as lexical diversity and syntactic complexity, but no statistically significant differences were found between grammatical accuracy and syntactic complexity.

The findings of this study are in line with those of Mazgutova and Kormos (2015), who found that as the proficiency level of students increases, the pieces of writing they produce become syntactically more complex and lexically more diverse. The findings also lend support to a previous study, which indicated that the three most predictive features of the level of proficiency are syntactic complexity, lexical diversity, and grammatical accuracy (McNamara et al., 2010). This conclusion makes sense with respect to theories of writing that propose that more skilled writers have a greater working memory capacity to access and use less familiar words as well as more complex syntax in their writing and more accurate grammar (Raimes, 2001).

However, the findings obtained from this study contradict those of some previous studies (e.g., Kim, 2014; Abrams, 2019). For example, Rahimi (2018) explored the effects of increasing the degree of reasoning and the number of elements on L2 argumentative writing. The findings revealed that by increasing task complexity, syntactic complexity was increased and had desirable effects on lexical diversity. These contradictory findings can be due to the amount of cognitive load (Sweller, 2011) applied by the learners in the task types. Generally, cognitive load refers to the total amount of mental activity imposed on working memory in any one instant (Bannert, 2002). In other words, cognitive load is strongly influenced by the number of elements in working memory that interact with each other (Sweller, 2011). Difficulty in writing is based on interacting elements that must be processed simultaneously (Kirkland & Saunders, 1991). Since

all the elements in writing cannot be done well at the same time, focus on one aspect causes a learner not to focus on other aspects (Skehan, 1998).

Furthermore, simultaneous increase in syntactic complexity and lexical diversity lends partial support to the cognition hypothesis (Robinson, 2011). Hamp-Lyons and Mathias (1994) argued that cognitively complex writing tasks more likely motivated learners to produce text types with rich syntactic complexity and lexical diversity than did cognitively less complex tasks. Accordingly, Gilabert (2007) concluded that simultaneous focus on syntactic complexity and lexical diversity is possible if tasks are manipulated to be cognitively complex along the resource-directing dimension.

Generally, in comparison with other studies (e.g., Abrams, 2019; Kim, 2014; Mazgutova & Kormos, 2015), when learners have freedom of choice in argumentative essays, the means of lexical diversity, grammatical accuracy, and syntactic complexity are higher than when freedom of choice is not given. However, in the present study, the means of grammatical accuracy and syntactic complexity were at the same level in comparison with lexical diversity which stood at higher level. This finding confirms that of Larsen-Freeman (2006), who also found the concurrent improvement of grammatical accuracy and syntactic complexity in argumentative writing. The study suggests that the high mean of lexically diverse items resulting from learners' freedom of choice comes at the cost of lower but concurrent rate of grammatical accuracy and complex syntax. The reason is, as far as argumentative essays are concerned, students at the time of their writing try to be as convincing as possible in order to outsmart the task or the potential reader (Johnson, 2017). Consequently, the primary focus on argumentation in argumentative essays leads to less attention to how accurately they write, as well as how complex their grammar is (Lahuerta, 2018), because post-hoc analyses in our study did not show any statistical significance between grammatical accuracy and syntactic complexity.

6. Conclusion and Implications

As the findings of the study show, as far as the freedom of choice of prompts is concerned, students tend to use a greater range of vocabulary and to be more accurate in the use of grammars with an increase in the level of syntactic complexity. Accordingly, it can be concluded that this improvement in those three dimensions may not depend on the types of essays, but it can be a direct result of the freedom given to the students to choose the topic of their interest. However, there might be some differences in the proportion of this improvement.

When learners are given freedom of choice, they are eager to use more syntactically complex structures, more lexically diverse items, and more accurate grammar while writing an argumentative essay. In other words, students are willing to use more of what they have learned when they are given more freedom in language classes. This is particularly true when it comes to students' use of lexically diverse items, which has received the highest score in argumentative essays. This high score in lexical diversity shows that language learners choose the topics that they are more comfortable with (Gebriel & Plakans, 2016), implying

that more comfort with the topic results in more diverse use of vocabulary. Thus, by involving students in decision making, they may be more eager to utilize what they have learned, irrespective of their argumentative performance.

It has to be noted that the results of this study are limited in several ways. These limitations necessitate future studies. The main limitation is perhaps the number of the participants. In this current investigation, the data were collected from only one university and one institute. The researchers could not collect more data from other educational centers. If more participants enroll in other research studies in similar contexts, the findings can be generalized and are more reliable. Another limitation is that the present study was a quantitative analysis using tests to explore if there is any significant difference between syntactic complexity, grammatical accuracy, and lexical diversity in Iranian EFL learners because it was not possible for the researchers to use observations or interviews. Future studies need to employ a mixed-methods research to come up with firm conclusions.



References

- Abdollahzadeh, E., Amini Farsani, M., & Beikmohammadi, M. (2017). Argumentative writing behavior of graduate EFL learners. *Argumentation* 31, 641–661.
- Abrams, Z. I. (2019). The effects of integrated writing on linguistic complexity in L2 writing and task-complexity. *System*, 81(1), 110-121.
- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92.
- Ahmadian, M. J., & Long, M. H. (Eds) (2022). *The Cambridge handbook of task-based language teaching*. Cambridge University Press.
- Ahmadian, M., & Mansouri A. (2020). Collaborative L1 planning and L2 written task performance in an Iranian EFL context. In L. Craig., & O Rhonda (Eds.) *Using tasks in diverse contexts* (pp. 237-255). Multilingual Matters.
- Allan, D. (2004). *Oxford placement test*. Oxford University Press.
- Allaw, E., & McDonough, K. (2019). The effect of task sequencing on second language written lexical complexity, accuracy, and fluency. *System*, 85. <https://doi.org/10.1016/j.system.2019.06.008>
- Bannert, M. (2002). Managing cognitive load—recent trends in cognitive load theory. *Learning and Instruction*, 12(1), 139-146.
- Bax, S. (2018). Text inspector [Computer Software]. Developed by Bax, S., <http://www.textinspector.com>
- Bayazidi, A., Ansarin, A. A., Mohammadnia, Z. (2019). The Relationship between syntactic and lexical complexity in speech monologues of EFL learners. *Applied Research on English Language*, 8(4), 473-488.
- Bean, J. C. (2011). *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom*. John Wiley & Sons.
- Beauvais, C., Olive, T., & Passerault, J. M. (2011). Why are some texts good and others not? Relationship between text quality and management of the writing processes. *Journal of Educational Psychology*, 103(2), 415.
- Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing*, 24(2), 183-202.
- Berman, R. A., & Nir-Sagiv, B. (2004). Linguistic indicators of inter-genre differentiation in later language development. *Journal of Child Language*, 31(2), 339-380.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23). John Benjamins.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.
- Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, 21(3), 487-508.
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, (pp. 23-46). John Benjamin.
- Chipere, N., Durán, P., Malvern, D., & Richards, B. J. (2004). *Lexical diversity and language development*. Palgrave Macmillan.
- Cho, M. (2019). The effects of prompts on L2 writing performance and engagement. *Foreign Language Annals*, 52, 576–594.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.
- Defazio, J., Jones, J., Tennant, F., & Hook, S. A. (2010). Academic literacy: The importance and impact of writing across the curriculum- A case study. *Journal of the Scholarship of Teaching and Learning*, 10(2), 34-47.

- Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference? In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 219-238). Amsterdam: John Benjamins.
- Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford University Press.
- Foster, P. & Skehan, P. (1999). The influence of planning on performance in task-based learning. *Paper presented at the British Association of Applied Linguistics, Leeds*.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375.
- Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing, 30*(4), 45-57.
- Gebril, A., & Plakans, L. (2016). Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency. *Journal of English for Academic Purposes, 24*(1), 78-88.
- Geranpayeh, A. (2006). A quick review of the English quick placement test. Retrieved from University of Cambridge ESOL Examinations: <http://www.uniss.it/documenti/lingue>.
- Gilbert, R. (2007). The simultaneous manipulation of task complexity along planning time and [+/-Here-and-Now]: Effects on L2 oral production. *Investigating Tasks in Formal Language Learning, 20*(1), 44-68.
- Grabe, W. (2002). *Genre in the classroom: Multiple perspectives*. Northern Arizona University
- Gray, C. D., & Kinnear, P. R. (2012). *IBM SPSS statistics 19 made simple*. Psychology Press.
- Guara-Tavares, M. G. (2008). *Pre-task planning. Working memory capacity and L2 speech performance* (Unpublished doctoral thesis). Univeridade Federal de Santa Catarina, Brazil.
- Hall, K. M., Sabey, B. L., & McClellan, M. (2005). Expository text comprehension: Helping primary-grade teachers use expository texts to full advantage. *Reading Psychology, 26*(3), 211-234.
- Hamp-Lyons, L. (2011). Writing assessment: Shifting issues, new tools, enduring questions. *Assessing Writing, 16*(1), 3-5.
- Hamp-Lyons, L., & Mathias, S. P. (1994). Examining expert judgments of task difficulty on essay tests. *Journal of Second Language Writing, 3*(1), 49-68.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473.
- Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing.
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. National Council of Teachers of English.
- Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English, 43*(7), 732-739.
- Ishikawa, T. (2007). The effect of manipulating task complexity along the [+/- Here-and- Now] dimension on L2 written narrative discourse. In M. D. P. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136-156). Multilingual Matters.
- Jabbari, A. A. (2014). Collocational differences in Persian and English and their effect on learners' production. *International Journal of Educational Investigations, 1*(2), 172-190.
- Jalilifar, A., Keyvan, M., & Don, A. (2017). The effect of writing prompts on developing argumentative essays of Iranian university students of English. *Journal of Modern languages, 27*, 48-76.
- Jeong, H. (2017). Narrative and expository genre effects on students, raters, and performance criteria. *Assessing Writing, 31*(1), 113-125.
- Jiang, W. (2013). Measurements of development in L2 written production: The case of L2 Chinese. *Applied Linguistics, 34*(1), 1-24.

- Johansson, V. (2009). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working Papers in Linguistics*, 53, 61-79.
- Johnson, M. D. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing*, 37(1), 13-38.
- Kamberelis, G., & Bovino, T. D. (1999). Cultural artifacts as scaffolds for genre development. *Reading Research Quarterly*, 34(2), 138-170.
- Kessler, M., Ma, W., & Solheim, I. (2021). The effects of topic familiarity on text quality, complexity, accuracy, and fluency: A conceptual replication. *TESOL Quarterly*. doi.org/10.1002/tesq.3096
- Kim, J. Y. (2014). Predicting L2 writing proficiency using linguistic complexity measures: A corpus-Based study. *English Teaching*, 69(4), 27-51.
- Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, 25(1), 105-121.
- Koutsoftas, A. D., & Gray, S. (2012). Comparison of narrative and expository writing in students with and without language-learning disabilities. *Language, Speech, and Hearing Services in Schools*, 19(2), 26-48.
- Kuiken, F., & Vedder, I. (2011). Task complexity and linguistic performance in L2 writing and speaking. *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 91-104). John Benjamins.
- Kuiken, F., & Vedder, I. (2012). Syntactic complexity, lexical variation, and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA*. (pp. 143-170). John Benjamins.
- Lahuerta, A. C. (2018). Study of accuracy and grammatical complexity in EFL writing. *International Journal of English Studies*, 18(1), 71-89.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied linguistics*, 27(4), 590-619.
- Liao, J. (2020). Do L2 lexical and syntactic accuracy develop in parallel? Accuracy development in L2 Chinese writing. *System*, 94, <https://doi.org/10.1016/j.system.2020.102325>
- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85-104.
- Martin-Beltrán, M., García, A.A., & Montoya-Avila, A. (2020). I know there's something like that in Spanish: Heritage language learners' multifaceted interactions with linguistically diverse peers. *International Journal of Applied Linguistics*, 30, 530-552.
- Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes Programme. *Journal of Second Language Writing*, 29(1), 3-15.
- McCarthy, M. (1998). *Spoken language and applied linguistics*. Ernst Klett Sprachen.
- McCarthy, P. M. & Jarvis, S. (2007). A theoretical and empirical evaluation of vocd, *Language Testing* 24(4), 459-88.
- McCarthy, P. M. (2005). An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). *Dissertation Abstracts International*, 66(12).
- McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381-392.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86.
- McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45(2), 499-515.

- Meihami, H., & Saadat, M. (2019). An investigation into the effects of prompt selection on writing complexity, accuracy, and fluency: The case of Iranian learners at different proficiency levels. *Journal of Language and Cultural Education*, 7(3), 79-96.
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47. <https://doi.org/10.1016/j.asw.2020.100511>.
- Newell, G. E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly*, 46(3), 273-304.
- Nippold, M. A. (2004). Research on later language development: International perspectives. *Language development across childhood and adolescence*, 3(1), 1-8.
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590-601.
- Polio, C. G. (2001). Research methodology in second language writing research: The case of text-based studies. *On second language writing* (pp. 91-115). Lawrence Erlbaum Associates.
- Rahimi, M. (2018). Effects of increasing the degree of reasoning and the number of elements on L2 argumentative writing. *Language Teaching Research*, 22(2), 1-22.
- Raimes, A. (2001). What unskilled ESL students do as they write: A classroom study of composing. In T. Silva & P. K. Matsuda (Eds.), *Landmark essays: On ESL writing* (pp. 37-62). Hermagoras Press.
- Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First Language*, 30(1), 3-26.
- Rebarber, T., Rybinski, P., Hauck, M., Scarcella, R., Buteux, A., & Wang, J. (2007). Designing the comprehensive English language learner assessment (CELLA) for the benefit of users. In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 63-79). University of California.
- Reilly, J., Zamora, A., & McGivern, R. F. (2005). Acquiring perspective in English: the development of stance. *Journal of Pragmatics*, 37(2), 185-208.
- Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis on language learning and performance* (pp. 3-37). John Benjamins.
- Robinson, P. (2015). The Cognition hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp.87-122). John Benjamins.
- Ruiz-Funes, M. (2015). Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing*, 28(2), 1-19.
- Rutherford, K. (2001). *An investigation of the effects of planning on oral production in a second language* [Unpublished MA thesis]. University of Auckland.
- Sadeghi, K., & Dilmaghani, S. K. (2013). The Relationship between lexical diversity and genre in Iranian EFL learners' writings. *Journal of Language Teaching and Research*, 4(2), 328-334.
- Scott, C. M. (1988). Spoken and written syntax. In M. Nippold (Ed.): *Later language development: Ages nine through nineteen* (pp. 49-55). Pro-Ed.
- Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, 43(2), 324-339.
- Scott, M. S., & Tucker, G. R. (1974). Error analysis and English-language strategies of Arab students. *Language Learning*, 24(1), 69-97.

- Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, 10(1), 209-232.
- Shi, B., Huang, L., & Lu, X. (2020). Effect of prompt type on test-takers' writing performance and writing strategy use in the continuation task. *Language Testing*, 37(3), 361-388.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge University Press.
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263-308.
- StataCorp, L. P. (2016). Stata MP for Windows 64-bit. *College Station, TX: StataCorp*.
- Sweller, J. (2011). Cognitive load theory. *Psychology of learning and motivation*. Academic Press.
- Tabachnick, B. G., & Fidell, L. S. (2014). Using multivariate statistics. Pearson Education Limited.
- Tavakoli, M., & Rezazadeh, M. (2014). Individual and collaborative planning conditions: Effects on fluency, complexity and accuracy in L2 argumentative writing. *Journal of Teaching Language Skills*, 32(4), 85-110.
- Templin, M. (1957). *Certain language skills in children: Their development and interrelationships*. University of Minnesota, The Institute of Child Welfare.
- Towell, R. (2012). Complexity, accuracy and fluency from the perspective of psycholinguistic second language acquisition research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 32-47). John Benjamins.
- Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, 39(3), 302-327.
- UCLES (University of Cambridge Local Examinations Syndicate) (2001). *Quick Placement Test*. Oxford University Press.
- Way, D. P., Joiner, E., & Seaman, M. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *The Modern Language Journal*, 84(2), 171-184.
- Weigle, S. C. (2002). *Assessing writing*. Ernst Klett Sprachen.
- Weigle, S. C. (2018). Scoring writing. In J. I. Liantas (Eds), *The TESOL encyclopedia of English language teaching* (pp.1-7). John Wiley. <https://doi.org/10.1002/9781118784235.eelt0546>.
- Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, 26(3), 445-466.
- Williams, S. (2018). Text inspector. *English Australia Journal*, 33(2), 80-83.
- Wolfe-Quintero, K., S. Ingaki, & H.-Y. Kim. (1998). *Second Language Development in Writing: Measures of fluency, accuracy, and complexity*. University of Hawai'i, Second Language Teaching and Curriculum Center.
- Zhang, X., & Lu, X. (2021). Revisiting the predictive power of traditional vs. fine-grained syntactic complexity indices for L2 writing quality: The case of two genres. *Assessing Writing*, 51. <https://doi.org/10.1016/j.asw.2021.100597>.